Using DASH Assisting Network Elements for Optimizing Video Streaming Quality

Jan Willem Kleinrouweler
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
j.w.m.kleinrouweler@cwi.nl

ABSTRACT

On-demand video streaming is a popular application which accounts for a large share of today's Internet traffic. Dynamic Adaptive Streaming over HTTP (DASH) is the major streaming technology used by large content providers. However, this technology suffers from performance problems when multiple clients are streaming on shared network links. Aiming at improving the viewers' Quality of Experience, this thesis studies how DASH Assisting Network Elements (DANEs) optimize bottleneck network links and improve DASH streaming performance. The DANE is aware of DASH traffic on the network link, and partitions available network resources between DASH players and other traffic. Two DANE prototypes have been proposed as part of this thesis. In experiments with DASH players in wired and wireless networks, it is shown that the DANEs increase video bitrate, reduce quality switches, and improve fairness between players. Additionally, Markov models have been developed to explore sharing policies for DANEs. The models are used to determine the effect of those policies on streaming performance and to optimize network resource sharing for DASH players. Thorough validations with real DASH players show that the Markov models are highly accurate. In the remaining last year of the PhD program, I will apply DANE technology to cellular (5G) networks and use the Markov models to obtain an optimal sharing strategy.

CCS CONCEPTS

• **Information systems** → **Multimedia streaming**; • **Networks** → *Network management*;

KEYWORDS

Dynamic Adaptive Streaming over HTTP (DASH); HTTP Adaptive Streaming; Video streaming; Performance modeling; Optimization

1 INTRODUCTION

Over-the-Top (OTT) video streaming is a popular Internet application. Content providers, such as YouTube[1] and Netflix[2], provide access to a large video catalog and serve many users every day. The popularity of VoD services marks a transition in media consumption: from traditional television broadcasting to on-demand video streaming over the Internet. With this transition, video delivery changes from using dedicated and managed infrastructures to HTTP-based streaming over the Internet. Dynamic Adaptive Streaming over HTTP (DASH) is the dominant technology for video streaming, and has been adopted by major content providers [23]. In DASH, a video is split into short segments (typically between two and ten seconds) and encoded into multiple bitrates and resolutions. The representations are listed in a manifest that is downloaded by the DASH player. DASH players select one of the representations based on the current network conditions. If network conditions change, DASH players can adapt by selecting another representations.

Using a simple HTTP-based protocol has the advantage that content providers can realize highly scalable delivery using Content Delivery Networks (CDNs). The disadvantage is that streaming issues arise, because DASH players have to adapt the video quality to match network conditions while downloading small video segments over HTTP/TCP [1, 2, 13]. This is challenging when streaming on shared network links. Performance issues result in freezes, low bitrate, quality switches, and unfairness. These issues do not match the viewers' expectations: playback should start immediately, playback should not be interrupted, video quality should be high, video quality should be constant [11]. Deterioration of the Quality of Experience (QoE) could eventually lead to disengagement and abandonment [8, 26], which may negatively affect the content providers (e.g. loss of revenue due to missed advertising) and the internet service providers (e.g. loss of customers after negative networking experiences).

To protect agains these streaming performance issues, and to deliver a high QoE to the viewers, it is important that all parts in the delivery chain are optimized for DASH. On the side of the content providers, optimization is done by caching and replicating video segments on CDNs. Bottlenecks are more often encountered at what is called "the last mile": the users' DSL connections, Wi-Fi networks, or cellular networks. These networks typically have limited bandwidth, but are often shared with multiple users. Networking is on best effort basis, which does not guarantee high throughputs for DASH. This leads to the central research question in this thesis:

How can last-mile network connections be optimized for DASH video streaming resulting in a better quality of experience for the viewers?

This thesis studies network elements, such as Wi-Fi routers or network gateways, that are made DASH aware and assist DASH players while streaming. These DASH Assisting Network Elements

[1]https://www.youtube.com/ (accessed May 28, 2017)
[2]https://www.netflix.com/ (accessed May 28, 2017)

MM '17, October 23–27, 2017, Mountain View, CA, USA

ACM ISBN 978-1-4503-4906-2/17/10.
https://doi.org/10.1145/3123266.3123965

(DANEs) divide the available network resources among DASH players and other traffic, and inform DASH players which bitrate they should select. The central research question is answered through the following related research questions:

- How can network assisted DASH improve the streaming performance in terms of video bitrate, number of quality switches, number of freezes, and fairness?
- How can network assisted DASH be realized, considering effectiveness and agility of the solution, as well as the privacy of the viewer?
- Which policies can be defined to divide network bandwidth among DASH players and other traffic?
- What is the effect of bandwidth sharing policies on DASH streaming performance?
- Can Markov modeling be used to efficiently evaluate bandwidth sharing policies in DANEs?

In the first three years of the PhD program, DANEs have been developed and evaluated in both wired and wireless networks. Prototypes show that DANEs effectively prevent freezes, increase the video bitrate, reduce quality switches, and improve fairness. With highly accurate Markov models, the effects of sharing policies in DANEs could be determined, and the models were used to optimize network resource sharing. In the remaining last year of the PhD program, I will apply DANE technology to cellular (5G) networks and use the Markov models presented in this thesis to obtain an optimal sharing strategy.

2 STATE OF THE ART

The goal of this thesis is to optimize the QoE of watching DASH based video streaming. We target DASH [27] because it is the major technology for online video streaming. DASH has been widely deployed [23] and accounts for a large share of today's Internet traffic [5]. Nevertheless, several studies have indicated performance issues with DASH in shared networks. Huang et al. show that DASH video quality decreases when it has to compete for resources with another TCP flow [13]. When network resources are shared with other DASH players, video quality becomes unstable (frequent changes in video quality), as demonstrated using off-the-shelf players in [1]. In [2], Akhshabi et al. identify the on-off download patterns of DASH players as the cause of instability. They recognize that bandwidth estimations in DASH players are a challenging task. This is confirmed by Esteban et al. in an analysis of TCP behavior with DASH [9]. DASH performance problems result in freezes, quality switches, and unfairness. DASH, in the way it is deployed now, does not yield a satisfiable QoE [8, 11, 26].

As an alternative to improving DASH adaptation algorithms, DASH Assisting Network Elements can enhance DASH performance. DANEs can roughly be classified in two categories: bandwidth reservation and bitrate signaling. In [12], the Houdaille and Gouache show that the stability of DASH streams can be improved with traffic shaping. Jarschel et al. use bandwidth reservation to increase the quality of YouTube streams using Software Defined Networking (SDN) tools [14]. Bitrate signaling using a proxy server based DANE is presented by Bouten et al. in [4]. Petrangeli et al. chain multiple proxy servers to address networks with multiple bottlenecks [24]. Georgopoulos use OpenFlow to discover DASH players and signal bitrates using a plugin in a DASH player [10]. Bandwidth reservation and bitrate signaling are combined by Cofano et al. in [6] and by Bentaleb et al. in [3]. This approach is similar to our implementation in [16]. Compared to the works mentioned before, we study in this thesis how different types of traffic control affect DASH adaptation algorithms and how much assistance has to be applied to be effective. Additionally, our implementation [16] has the advantages to be lightweight, privacy friendly (i.e. it does not use deep packet inspection), and it fully supports encrypted streams over HTTPS.

Currently, communication between DANE and DASH player is being standardized in the Server and Network Assisted DASH (SAND) proposal [29]. This proposal specifies the message format, but not how the messages should be handled in player, nor how the DANE should divide the available bandwidth. We attempt to provide a tool for optimizing bandwidth sharing in the form of Markov models. To the best of our knowledge, no related work exists on modeling sharing policies in DANEs. DASH player behavior is modeled by Tanwir and Perros in [28]. However the authors modeled DASH from the perspective of a single DASH player, and it is thus not suitable for optimizing sharing policies in DANEs. The models presented in this thesis are targeting that use case.

3 DASH ASSISTING NETWORK ELEMENT

This thesis studies DANEs that overcome DASH performance issues. The two problems that are the source for DASH performance issues are: errors in bandwidth estimations done by DASH players [1] and limited network throughput due the use of HTTP/TCP [13]. In (plain) DASH, the player determines the video quality (i.e. bitrate and resolution) of the stream. A DASH player bases the quality of a next video segment based on the download speeds for previous segments and the current buffer level. Bandwidth estimations are difficult for DASH players due to the limited view on network activity. TCP with DASH fails to reach the required high throughputs because of the separated segment downloads. Network throughput is limited to TCP, and cannot be influenced by a DASH player. Solely improving DASH adaptation algorithms will thus not overcome the performance problems. DANEs, being network elements (e.g. Wi-Fi routers or network gateways), have a better overview of network capabilities and current network traffic compared to individual DASH players. They make well-informed decisions on how available network resources are shared among DASH players and other traffic.

We consider a DANE to be network element that is in the path between the DASH player and an HTTP server, and is preferably located close to the bottleneck link on the last-mile. High level functioning of this DANE is illustrated in Figure 1. DASH players first report their existence to the DANE (step 1). They send the representations (i.e. bitrates) from the DASH manifest together with characteristics of the device. At this point, the DANE has an overview of active DASH players. Combining it with statistics of recent other network traffic, it divides the available network resources and assigns each DASH player a target representation. The target bitrate is communicated to the DASH players (step 2). The DASH players use the target bitrate when download speeds are sufficient. To ensure that DASH players are not stuck at low

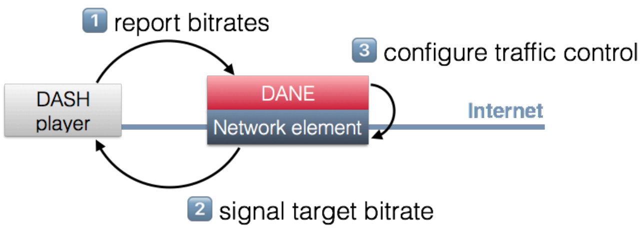

Figure 1: DANE flow of events (Illustration from [15])

throughputs, the DANE configures traffic control on the network element (step 3).

For this thesis, two DANE prototypes have been proposed. The first is a proxy server based DANE that is presented in [18]. HTTP traffic is routed through the DANE where it is inspected to detect DASH flows. Stream characteristics are obtained by reading the manifest file. Adaptation assistance is provided by altering the contents in the HTTP request header. Our results show that this DANE greatly improves the stability (bitrate) and fairness compared to DASH players without DANE. An improved version of the DANE is presented in [16]. This DANE works out-of-band via a WebSocket connection from DASH players to DANE. The out-of-band DANE has the following advantages over the proxy server based DANE: (1) it is more lightweight, HTTP traffic does not have to be inspected reducing the computational load on the DANE; (2) it improves the users' privacy, only minimal information is shared with the DANE (i.e. available video representations, and optionally device characteristics), which stream a user is watching stays private; and (3) it is compatible with encrypted streams over HTTPS, the connection from DASH player to server is not terminated (as with the proxy server), maintaining the chain of trust. Extensive evaluations in a Wi-Fi network show that this DANE outperforms state-of-art DASH adaptation algorithms [16, 17]. In a crowded network our DANE effectively eliminates freezes, increases video bitrate, reduces quality switches, and establishes fairness (as in bitrate). In [22], the scalability of the out-of-band DANE is enhanced. To be able to cope with a large number of frequently starting and stopping DASH players, the transactions between DANE and DASH players, as well as between DANE and network switch are reduced. The DANE is evaluated in a testbed with Raspberry Pis, hosting (up to) 600 DASH players.

4 DANE PERFORMANCE MODELING

Compared to plain DASH, where the players determine the streaming quality, DANEs (partially) take over this task from the players. The DANE has an overview of active DASH players and other network traffic, which has the advantage that it is better informed while making decisions. A simple sharing policy in a DANE would divide the bandwidth equally over the active DASH players. However, networks are often shared by devices that have different characteristics and requirements (e.g. screens size, resolution, priority). A different sharing policy could potentially better address the diversity of devices. The sharing policy, combined with the mix of devices, how often they start and stop, and the network capabilities, determines the final streaming quality. Simulations or experiments in testbeds are often used to evaluate a sharing policy. This approach is unpractical when many policies have to be evaluated, because changing parameters requires the simulations or experimental runs to be repeated. Therefore, this thesis investigates if Markov modeling can replace simulations and testbed experimentations to speed up evaluations, as well as be used for optimizing sharing policies.

When using a DANE, DASH players stream at the target bitrates that they receive from the DANE most of the time. This makes streaming bitrates predictable given the sharing policy and number of active players, and thus potentially interesting for modeling. For the first version of the model in [18], we modeled starting and stopping DASH players as birth-death Markov process. Different types of players (e.g. device type or priority) are included in this model by letting it grow in multiple dimensions. Each state in the process represents a number of active players of each type. For each state in the Markov process, the sharing policy is applied, meaning that for each combination of players, the streaming bitrates are computed. The Markov model was then analytically solved, resulting in observations of how much time is spend in each state and the frequencies of how often the process transitions between states. Based on those observations, the mean video bitrate and the number of quality switches (these two factors play an important role in the viewers' QoE [7, 11, 25]) were obtained for each device type.

In extensive validations in [18, 21], that compare our model with DASH performance using DANEs, we show that our model is highly accurate. Only when DASH players use large buffers, the model underestimates the expected streaming bitrate. Since commercial players often use large buffer (at least 30 seconds), we improved our model to be sensitive to buffer sizes [21]. The model was applied in [20] to compare different sharing policies for DANEs: DASH priority, background traffic priority, and an in-between policy. The in-between policy balances the bandwidth between DASH- and background traffic. A weighted-sum optimization function finds the parameters for this policy. This resulted in a policy that reduced the number of quality switches by 55%, left the average video bitrate unchanged, and reduced the probability that there was not enough bandwidth allocated to background traffic to 17%.

5 STATUS & OUTLOOK

Efforts in the first three years of the PhD program have led to publications in international conferences (i.e. [16, 18, 20, 22]) and journals (i.e [17, 19, 21]). A timeline overview of the contributions with corresponding papers is given in Figure 2. Contributions have been made in the areas of multimedia systems and performance modeling: this thesis studies technical aspects of DANEs as well as abstract analytical performance models to optimize DANE policies. Nevertheless, work in these topics is centered around DANEs and this thesis successfully joins these two areas.

Up to this point, the focus has been on DANEs in wired and Wi-Fi networks. In general, the total capacity of those networks stays constant, each device can potentially use the full capacity, and the bandwidth can be divided using traffic control tools such as Linux tc[3]. However, a significant part of video streaming is consumed using mobile (cellular) networks [5]. Current efforts are focused

[3] http://lartc.org/manpages/tc.txt (accessed May 28, 2017)

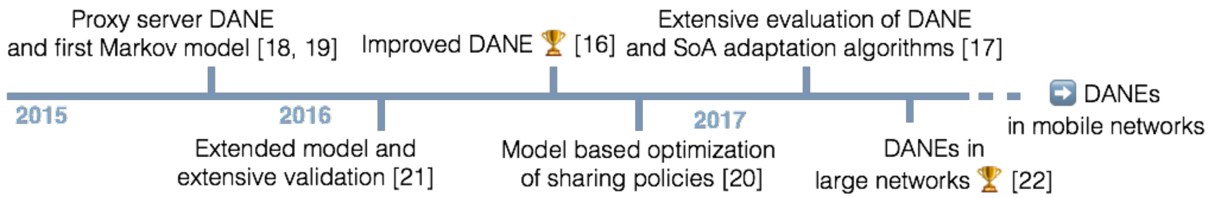

Figure 2: Timeline overview of publications in the thesis. For [16] and [22] we received an 'Excellence in DASH' best paper award, sponsored by the DASH Industry Forum.

on moving DASH network assistance to mobile networks. In mobile networks, the network throughput to a device is influenced by the channel quality. Depending on the channel quality, a different coding with more or less redundancy is used. In addition, the scheduler in the mobile access point can assign devices a number of resources (resource blocks). Current mobile networks are configured to schedule resources independent of the application, leading to an unsatisfiable experience for DASH. In the light of 5G, we are investigating if scheduling of a network slice dedicated to video streaming, can be optimized when it is fed with state information (desired video quality, buffer levels, etc.) from DASH players. The Markov model in this thesis is generally applicable and will be used for optimizing the resource scheduler. Realizing this optimization requires interactions between network and DASH players similar to the DANEs that have been studied in this thesis. Results from the optimization, the target video quality and buffering mode, are communicated back to the DASH players. The goal of this DANE is to more efficiently divide network resources, such that either the video quality can be increased, or more DASH players can be served in the network.

Looking further ahead, DANEs have the potential to be an enabler of novel multimedia applications. Applications such as Virtual/Augmented Reality (VR/AR), multiscreen immersive TV, and 360 degree video, require instant available of high network bandwidth. In this sense, it makes these applications comparable to video on demand. However, user interactivity results in tighter system requirements: the DANE must provide the assistance and resources without delay. This posing challenging issues for future research. At the ACM MM Doctoral Symposium, I would like to seek the advise of the MultiMedia community, and further explore the possibilities of network assistance to enhance the experience of upcoming multimedia applications.

ACKNOWLEDGEMENTS

Jan Willem is a PhD student at Centrum Wiskunde & Informatica, supervised by Pablo Cesar, Prof. Rob van der Mei, and Prof. Dick Bulterman.

REFERENCES

[1] Saamer Akhshabi, Lakshmi Anantakrishnan, Ali C Begen, and Constantine Dovrolis. 2012. What happens when HTTP adaptive streaming players compete for bandwidth?. In *NOSSDAV '12: Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video*. New York, New York, USA, 9–14.

[2] Saamer Akhshabi, Ali C Begen, and Constantine Dovrolis. 2011. An Experimental Evaluation of Rate-adaptation Algorithms in Adaptive Streaming over HTTP. In *Proceedings of the Second Annual ACM Conference on Multimedia Systems*. ACM, New York, NY, USA, 157–168.

[3] Abdelhak Bentaleb, Ali C. Begen, and Roger Zimmermann. 2016. SDNDASH: Improving QoE of HTTP Adaptive Streaming Using Software Defined Networking. In *Proceedings of the 2016 ACM on Multimedia Conference (MM '16)*. ACM, New York, NY, USA, 1296–1305.

[4] N. Bouten, J. Famaey, S. LatrÃĪ, R. Huysegems, B. D. Vleeschauwer, W. V. Leekwijck, and F. D. Turck. 2012. QoE optimization through in-network quality adaptation for HTTP Adaptive Streaming. In *2012 8th international conference on network and service management (cnsm) and 2012 workshop on systems virtualiztion management (svm)*. 336–342.

[5] Cisco. 2017. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016âĂŞ2021 White Paper. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html (accessed May 29, 2017). (2017).

[6] G. Cofano, L. De Cicco, T. Zinner, A. Nguyen-Ngoc, P. Tran-Gia, and S. Mascolo. 2016. Design and Experimental Evaluation of Network-assisted Strategies for HTTP Adaptive Streaming. In *Proceedings of the 7th International Conference on Multimedia Systems (MMSys '16)*. ACM, New York, NY, USA, Article 3, 12 pages.

[7] Nicola Cranley, Philip Perry, and Liam Murphy. 2006. User perception of adapting video quality. *International Journal of Human-Computer Studies* 64, 8 (2006), 637–647.

[8] Florin Dobrian, Asad Awan, Dilip Joseph, Aditya Ganjam, Jibin Zhan, Vyas Sekar, Ion Stoica, and Hui Zhang. 2013. Understanding the Impact of Video Quality on User Engagement. *Commun. ACM* 56, 3 (March 2013), 91–99.

[9] Jairo Esteban, Steven A. Benno, Andre Beck, Yang Guo, Volker Hilt, and Ivica Rimac. 2012. Interactions Between HTTP Adaptive Streaming and TCP. In *Proceedings of the 22Nd International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '12)*. ACM, New York, NY, USA, 21–26.

[10] Panagiotis Georgopoulos, Yehia Elkhatib, Matthew Broadbent, Mu Mu, and Nicholas Race. 2013. Towards network-wide QoE fairness using openflow-assisted adaptive video streaming. In *FhMN '13: Proceedings of the 2013 ACM SIGCOMM workshop on Future human-centric multimedia networking*. ACM Request Permissions, New York, NY, USA, 15–20.

[11] Tobias Hoßfeld, Michael Seufert, Christian Sieber, Thomas Zinner, and Phuoc Tran-Gia. 2015. Identifying QoE optimal adaptation of HTTP adaptive streaming based on subjective studies. *Computer Networks* 81 (2015), 320–332.

[12] Rémi Houdaille and Stéphane Gouache. 2012. Shaping HTTP adaptive streams for a better user experience. In *MMSys '12: Proceedings of the 3rd Multimedia Systems Conference*. ACM Request Permissions, New York, New York, USA, 1–9.

[13] Te-Yuan Huang, Nikhil Handigol, Brandon Heller, Nick McKeown, and Ramesh Johari. 2012. Confused, timid, and unstable: picking a video streaming rate is hard. In *IMC '12: Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM Request Permissions, New York, New York, USA, 225–238.

[14] M. Jarschel, F. Wamser, T. Hohn, T. Zinner, and P. Tran-Gia. 2013. SDN-Based Application-Aware Networking on the Example of YouTube Video Streaming. In *2013 Second European Workshop on Software Defined Networks*. 87–92.

[15] Jan Willem Kleinrouweler. 2017. Enhancing over-the-top video streaming quality with DASH assisting network elements. In *Adjunct Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '17 Adjunct)*. New York, NY, USA, 4 pages.

[16] Jan Willem Kleinrouweler, Sergio Cabrero, and Pablo Cesar. 2016. Delivering Stable High-quality Video: An SDN Architecture with DASH Assisting Network Elements. In *Proceedings of the 7th International Conference on Multimedia Systems (MMSys '16)*. ACM, New York, NY, USA, Article 4, 10 pages.

[17] Jan Willem Kleinrouweler, Sergio Cabrero, and Pablo Cesar. 2017. An SDN Architecture for Privacy-friendly Network Assisted DASH. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 3s, Article 44 (2017), 22 pages.

[18] Jan Willem Kleinrouweler, Sergio Cabrero, Rob van der Mei, and Pablo Cesar. 2015. Modeling Stability and Bitrate of Network-Assisted HTTP Adaptive Streaming Players. In *27th International Teletraffic Congress (ITC 27)*. Ghent, Belgium.

[19] Jan Willem Kleinrouweler, Sergio Cabrero, Rob van der Mei, and Pablo Cesar. 2015. Modeling the Effect of Sharing Policies for Network-assisted HTTP Adaptive Video Streaming. *SIGMETRICS Performance Evaluation Review* 43, 2 (Sept. 2015), 26–27.

[20] Jan Willem Kleinrouweler, Sergio Cabrero, Rob van der Mei, and Pablo Cesar. 2016. A Markov Model for Evaluating Resource Sharing Policies for DASH Assisting Network Elements. In *28th International Teletraffic Congress (ITC 28)*. Ghent, Belgium.

[21] Jan Willem Kleinrouweler, Sergio Cabrero, Rob van der Mei, and Pablo Cesar. 2016. A model for evaluating sharing policies for network-assisted HTTP adaptive streaming. *Computer Networks* 109, Part 2 (2016), 234 – 245. Traffic and Performance in the Big Data Era.

[22] Jan Willem Kleinrouwer, Britta Meixner, and Pablo Cesar. 2017. Improving Video Quality in Crowded Network Using a DANE. In *Proceedings of the 27th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '16)*. ACM, New York, NY, USA, 6 pages.

[23] Stefan Lederer. 2015. Why YouTube & Netflix use MPEG-DASH in HTML5. Availble online https://bitmovin.com/status-mpeg-dash-today-youtube-netflix-use-html5-beyond/ (accessed February 8, 2017). (Februari 2015).

[24] Stefano Petrangeli, Jeroen Famaey, Maxim Claeys, Steven Latré, and Filip De Turck. 2015. QoE-Driven Rate Adaptation Heuristic for Fair Adaptive Video Streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 2 (Oct. 2015), 28:1–28:24.

[25] Michael Seufert, Tobias Hosfeld, and Christian Sieber. 2015. Impact of intermediate layer on quality of experience of HTTP adaptive streaming. In *2015 11th International Conference on Network and Service Management (CNSM)*. IEEE, 256–260.

[26] R. K. Sitaraman. 2013. Network performance: Does it really matter to users and by how much?. In *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)*. 1–10.

[27] I Sodagar. 2011. The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *Industry and Standards* (2011).

[28] Savera Tanwir and Harry Perros. 2016. Modeling live adaptive streaming over HTTP. *Computer Communications* 85 (2016), 74 – 88.

[29] E Thomas, M O van Deventer, T Stockhammer, A C Begen, and J Famaey. 2015. Enhancing MPEG dash performance via server and network assistance. In *The Best of IET and IBC*. Institution of Engineering and Technology, 48–53.